14

CLAIMS

1. Data classification apparatus comprising:

    an input device for receiving a plurality of training classified examples and at least one unclassified example;

    a memory for storing the classified and unclassified examples;

    an output terminal for outputting a predicted classification for the at least one unclassified example; and

    a processor for identifying the predicted classification of the at least one unclassified example

    wherein the processor includes:

    classification allocation means for allocating potential classifications to each unclassified example and for generating a plurality of classification sets, each classification set containing the plurality of training classified examples and the at least one unclassified example with its allocated potential classification;

    assay means for determining a strangeness value valid under the iid assumption for each classification set;

    a comparative device for selecting the classification set to which the most likely allocated potential classification for the at least one unclassified example belongs, wherein the predicted classification output by the output terminal is the most likely allocated classification according to the strangeness values assigned by the assay means; and

a strength of prediction monitoring device for determining a confidence value for the predicted classification on the basis of the strangeness value assigned by the assay means to one of the classification

5   sets to which the second most likely allocated potential classification of the at least one unclassified example belongs.

2.   Data classification apparatus as claimed in claim 1,

10   wherein the processor further includes an example valuation device which determines individual strangeness values for each training classified example and the at least one unclassified example having an allocated potential classification.

15

3.   Data classification apparatus as claimed in claim 2, wherein Lagrange multipliers are used to determine the individual strangeness value.

20 4.   Data classification apparatus as claimed in claim 2, wherein the assay means determines a strangeness value for each classification set in dependence on the individual strangeness values of each example.

25 5.   Data classification apparatus comprising:
       an input device for receiving a plurality of training classified examples and at least one unclassified example;
       a memory for storing the classified and unclassified

30   examples;

16

stored programs including an example classification program;

an output terminal for outputting a predicted classification for the at least one unclassified example; and

5

a processor controlled by the stored programs for identifying the predicted classification of the at least one unclassified example wherein the processor includes:

classification allocation means for allocating potential classifications to each unclassified example and for

10

generating a plurality of classification sets, each classification set containing the plurality of training classified examples and the at least one unclassified example with its allocated potential classification;

assay means for determining a strangeness value valid

15

under the iid assumption for each classification set;

a comparative device for selecting the classification set to which the most likely allocated potential classification for the at least one unclassified example belongs, wherein

the predicted classification output by the output

20

terminal is the most likely allocated potential classification according to the strangeness values assigned by the assay means and

a strength of prediction monitoring device for determining a confidence value for the predicted

25

classification on the basis of the strangeness value assigned by the assay means to one of the classification sets to which the second most likely allocated potential classification of the at least one unclassified example

30

belongs.

17

6. A data classification method comprising:

inputting a plurality of training classified examples and at least one unclassified example;

identifying a predicted classification of the at least one

5 unclassified example which includes,

allocating potential classifications to each unclassified example;

generating a plurality of classification sets, each classification set containing the plurality of training

10 classified examples and the at least one unclassified example with its allocated potential classification;

determining a strangeness value valid under the iid assumption for each classification set;

selecting the classification set to which the most likely

15 allocated potential classification for the at least one unclassified example belongs, wherein the predicted classification is the most likely allocated potential classification in dependence on the strangeness values;

determining a confidence value for the predicted

20 classification on the basis of the strangeness value assigned to one of the classification sets to which the second most likely allocated potential classification for the at least one unclassified example belongs; and

outputting the predicted classification for the at least

25 one unclassified example and the confidence value for the predicted classification.

7. A data classification method as claimed in claim 6, further including determining individual strangeness

30 values for each training classified example and the at

18

least one unclassified example having an allocated potential classification.

8. A data classification method as claimed in any one of the
5 preceding claims, wherein the selected classification set is selected without the application of any general rules determined from the training set.

9. A data carrier on which is stored a classification program
10 for classifying data by performing the following steps:
generating a plurality of classification sets, each classification set containing a plurality of training classified examples and at least one unclassified example that has been allocated a potential classification;
15 determining a strangeness value valid under the iid assumption for each classification set;
selecting the classification set to which the most likely allocated potential classification for the at least one unclassified example belongs, wherein the predicted
20 classification is the most likely allocated potential classification in dependence on the strangeness values; and
determining a confidence value for the predicted classification on the basis of the strangeness value
25 assigned to one of the classification sets to which the second most likely allocated potential classification for the at least one unclassified example belongs.

30

AMENDED SHEET